Neural Abstractive Text Summarization and Fake News Detection

Soheil Esmaeilzadeh (soes@stanford.edu), Gao Xian Peh (gxpeh@stanford.edu), Angela Xu (aaxu@stanford.edu)

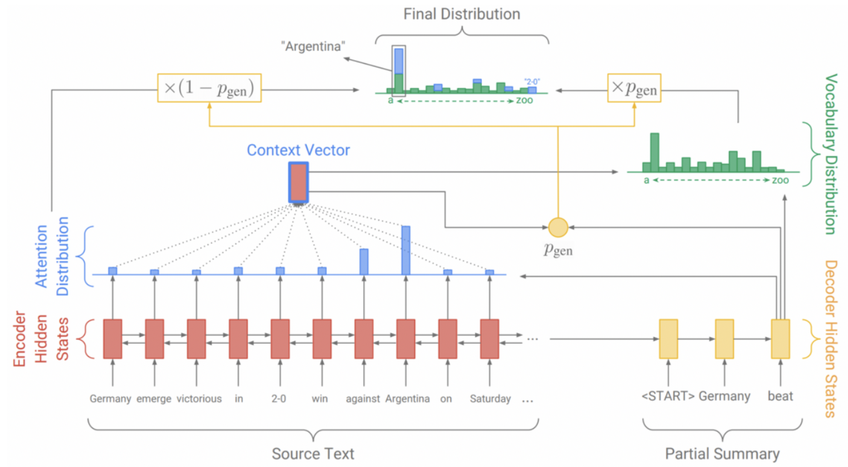
Motivation

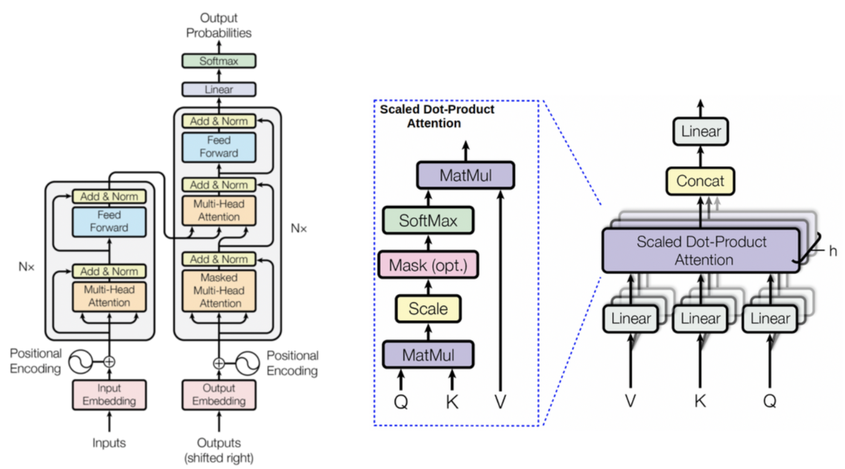
In this work, we study **abstractive text summarization** by exploring and comparing models such as LSTM-encoder-decoder + attention (baseline), pointer-generator + coverage, and transformers. As an extension of our work, we apply our text summarization model as a **feature extractor** for a **fake news detection** task where the news articles prior to classification are summarized and the results are compared against the classification using either the original news text or the headline text.

Datasets

Summarization dataset: For summarization task we use the CNN-Dailymail dataset provided by DeepMind with a split of 287,200 (92%), 13,360 (4.2%), 11,400 (3.8%) as train, dev, and test sets respectively.
Fake news dataset: For fake news classification task, we use a fake news dataset with headlines and article content provided by George McIntire with a split of 80%, 10%, 10% as train, dev, and test sets respectively. The dataset contains 3164 fake news and 3171 real articles.

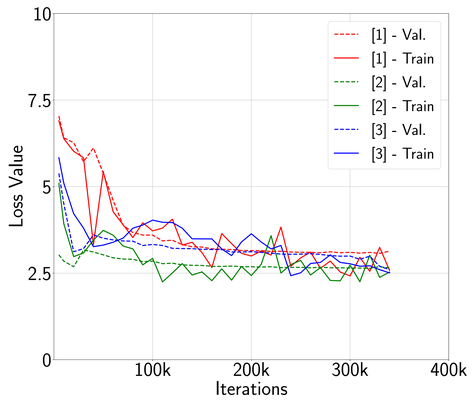
Models





Top: Pointer-generator network [See et al.], **Bottom**: Transformers architecture [Kiyoharu et al.]

Text Summarization Results



Validation and training loss values v.s. the number of iterations for the summarization models (19 epochs)

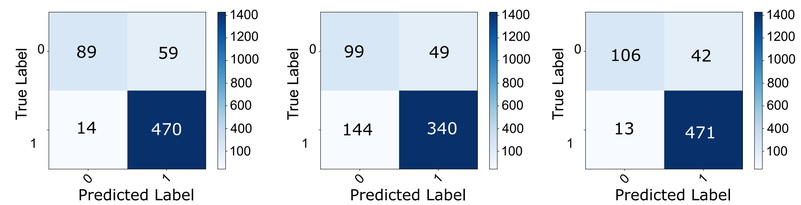
Model	ROUGE F1 Scores		
	1	2	L
[1]	35.68	14.21	30.56
[2]	38.47	16.33	33.37
[3]	38.97	16.81	35.41
[4]	36.55	15.21	31.19

[1]. LSTM encoder decoder + attention (baseline), [2]. baseline + pointer-generator, [3]. baseline + pointer-generator + coverage, [4]. transformers

Reference	once a super typhoon , maysak is now a tropical storm with 70 mph winds . it could still cause flooding , landslides and other problems in the philippines .
Model [1]	[UNK] gained super typhoon status thanks to its sustained 150 mph winds. it 's now classified as a tropical storm. it 's expected to make landfall sunday on the southeastern coast of [UNK] province .
Model [2]	tropical storm maysak approached the asian island nation saturday . it's now classified as a tropical storm , according to the philippine national weather service . it's now classified as a tropical storm , according to the philippine weather service .
Model [3]	just a few days ago , maysak gained super typhoon status thanks to its sustained 150 mph winds . it 's now classified as a tropical storm , according to the philippine national weather service .
Model [4]	super typhoon could weaken . new jersey , but it will . philippine ocean strength . at least 132 people are injured , including 18 .

Comparison of the generated summary using the summarization models v.s. the ground truth

Fake News Detection Results



Full body text | Headline text | Summarized text

Input Features	Cell Type & Size	Dropout	Accuracy %	Average Length (in words)
Full body text	LSTM-64	0.2	92	10.51
Headline text	Bi-LSTM-64	0.2	91	387.89
Summary text	Bi-LSTM-128	0.2	**93**	20.41

Fake news classifier results

References

[1] Abigail See, Peter J Liu, and Christopher D Manning. Get To The Point: Summarization with Pointer-GeneratorNetworks. 2017. ISSN 15420752. doi: 10.18653/v1/P17-1099.
[2] Kiyoharu J. Miyagishima, Ulrike Grünert, and Wei Li. Processing of S-cone signals in the inner plexiform layerof the mammalian retina, 2014. ISSN 14698714.